

UC Davis

UC Davis Previously Published Works

Title

Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics.

Permalink

<https://escholarship.org/uc/item/9mw0554m>

Journal

Metabolites, 8(2)

ISSN

2218-1989

Authors

Blaženović, Ivana
Kind, Tobias
Ji, Jian
et al.

Publication Date

2018-05-01


DOI

10.3390/metabo8020031

Peer reviewed

Review

Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics

Ivana Blaženović¹, Tobias Kind¹, Jian Ji² and Oliver Fiehn^{1,3,*} 

¹ NIH West Coast Metabolomics Center, UC Davis Genome Center, University of California, Davis, CA 95616, USA; iblazenovic@ucdavis.edu (I.B.); tkind@ucdavis.edu (T.K.)

² State Key Laboratory of Food Science and Technology, School of Food Science of Jiangnan University, School of Food Science Synergetic Innovation Center of Food Safety and Nutrition, Wuxi 214122, China; jijianjndx@126.com

³ Department of Biochemistry, Faculty of Sciences, King Abdulaziz University, Jeddah 21589, Saudi Arabia

* Correspondence: ofiehn@ucdavis.edu; Tel.: +1-530-752-9922

Received: 8 April 2018; Accepted: 6 May 2018; Published: 10 May 2018



Abstract: The annotation of small molecules remains a major challenge in untargeted mass spectrometry-based metabolomics. We here critically discuss structured elucidation approaches and software that are designed to help during the annotation of unknown compounds. Only by elucidating unknown metabolites first is it possible to biologically interpret complex systems, to map compounds to pathways and to create reliable predictive metabolic models for translational and clinical research. These strategies include the construction and quality of tandem mass spectral databases such as the coalition of MassBank repositories and investigations of MS/MS matching confidence. We present in silico fragmentation tools such as MS-FINDER, CFM-ID, MetFrag, ChemDistiller and CSI:FingerID that can annotate compounds from existing structure databases and that have been used in the CASMI (critical assessment of small molecule identification) contests. Furthermore, the use of retention time models from liquid chromatography and the utility of collision cross-section modelling from ion mobility experiments are covered. Workflows and published examples of successfully annotated unknown compounds are included.

Keywords: tandem mass spectrometry; library search; in silico fragmentation; high resolution mass spectrometry; compound identification; metabolomics

1. Introduction

Metabolomics is the comprehensive study of small molecules present in cells, tissues and body fluids. Advances in metabolic profiling have led to discoveries of biomarkers in a variety of medical conditions using metabolomics and lipidomics approaches, including the vision to utilize metabolomics for precision medicine [1–3]. Untargeted metabolomics experiments allow for the acquisition of thousands of metabolite signals in a single sample [4]. However, a large percentage of these signals remain structurally unknown [5], and therefore compound identification remains one of the large obstacles in metabolomics [6,7].

Currently, two major analytical platforms are used in the small molecule identification process. Nuclear magnetic resonance (NMR) is a powerful structure elucidation technique and it has a significant advantage due to its nondestructive and noninvasive characteristics of analysis. However, this method lacks the sensitivity needed for the simultaneous analysis of thousands of metabolites observed in biological samples [8,9]. High resolution chromatographic separation techniques coupled to accurate tandem mass spectrometry (LC-MS/MS) represents the most important metabolomics platform. This technology allows for the physical separation of thousands of metabolites and therefore provides a more comprehensive view of the metabolome.

Classical structure elucidation using NMR commonly elucidates the full structure using de-novo approaches [10]. The natural product [11], environmental [12] and mass spectrometry community [13] usually have different definitions for compound identification. In metabolomics, five different levels exist (see Table 1) including the new ‘Level 0’ that requires the full 3D structure and stereochemistry information. More common are ‘Level 1’ annotations that are confirmed by two orthogonal parameters, such as retention time and MS/MS spectrum. These levels were initially forged by the Metabolomics Standards Initiative (MSI) of the Metabolomics Society [14,15] and were later refined by the compound identification workgroup of the society. It is recommended to integrate the level of annotation for each compound into metabolomic profiling reports.

Table 1. New confidence levels of compound annotations, as discussed by the Compound Identification work group of the Metabolomics Society at the 2017 annual meeting of the Metabolomics Society (Brisbane, Australia). The new addition refers to the ‘Level 0’ annotation; other levels remain as discussed by the Metabolomics Standards Initiative.

Confidence Level	Description	Minimum Data Requirements
Level 0	Unambiguous 3D structure: Isolated, pure compound, including full stereochemistry	Following natural product guidelines, determination of 3D structure
Level 1	Confident 2D structure: uses reference standard match or full 2D structure elucidation	At least two orthogonal techniques defining 2D structure confidently, such as MS/MS and RT or CCS
Level 2	Probable structure: matched to literature data or databases by diagnostic evidence	At least two orthogonal pieces of information, including evidence that excludes all other candidates
Level 3	Possible structure or class: Most likely structure, isomers possible, substance class or substructure match	One or several candidates possible, requires at least one piece of information supporting the proposed candidate
Level 4	Unknown feature of interest:	Presence in sample

A number of reviews have been published that cover many diverse metabolomics topics including chromatography, data processing and statistics in great detail [16–24]. We mostly focus on papers that discuss structure elucidation approaches involving liquid chromatography tandem mass spectrometry (LC-MS/MS) within the last 5–10 years. The review is thematically divided into important sections that include mass spectral database search, in silico fragmentation tools and orthogonal coupled techniques including retention time matching and ion mobility spectrometry (see Figure 1). Lipidomics and mass spectral imaging approaches are not fully covered. Classical chemical derivatization and isotope labeling studies are discussed elsewhere [25]. Here, we only discuss a selected number of software tools and databases than can help practitioners to obtain results during the annotation of unknown compounds; larger surveys were covered in [17,23,26].

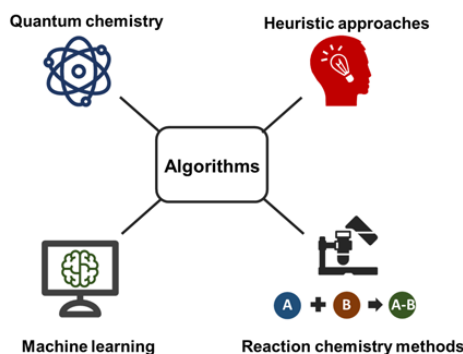


Figure 1. Computational metabolomics approaches help to unravel the complexity of the metabolome and especially shed light on unknown metabolites. This includes technologies across different disciplines, including quantum chemistry, machine learning, heuristic approaches and reaction chemistry-based methods.

2. Compound Databases and Chemical Space

The chemical space of small molecules currently covered in databases such as PubChem, ChemSpider or the Chemical Abstracts Database is larger than 120 million compounds [16] (see Table 2). The number of compounds with biological relevance is estimated at 1–2 million [27]. However, a large majority of metabolites discovered during untargeted metabolic profiling remains unknown, including many microbial [28], environmental [29] and natural compounds. In fact, very few reports in published research have more than 20% identified compounds in untargeted analysis, as can be seen at the Metabolomics Workbench [30], or the European metabolomics repository MetaboLights [31].

Table 2. Overview of selected compound databases commonly used for compound identification.

Database	Targets	Description
PubChem [32]	All small molecules	Small molecules, metadata
ChemSpider [33]	All small molecules	Small molecules, curated data
KEGG [34]	Metabolites	Pathway database, multiple species
MetaCyc [35]	Metabolites	Pathway database, multiple species
BRENDA [36]	Enzymes	Enzyme and metabolism data
HMDB [37]	Metabolites	Human metabolites
CHEBI [38]	Small molecules	Molecules of biological interest
UNPD [39]	Metabolites	Secondary plant metabolites
MINE [40]	Metabolites	In silico predicted metabolites

During the structure elucidation process, small molecule databases serve as a foundation of known and well-researched metabolites (see Table 2). Enzyme and pathway databases such as KEGG, MetaCyc and BRENDA serve as connectors to the proteomics and transcriptomics domain. Molecular formulas or accurate masses can be queried in such databases, and potential structure candidates can be retrieved to be investigated by in-silico fragmentation software tools. In many cases, it is important to restrict the search space by including taxonomy information. Molecular discovery in humans can be obtained from the Human Metabolome Database (HMDB) [37], and plant researchers should restrict their search space to primary and secondary plant metabolites such as found in the UNPD (Universal Natural Product Database) database [39] or compounds covered in the natural product space [41,42]. For exposome related research, environmental database resources can be utilized [43,44].

In case the compounds have not yet been described in the literature, enzymatic expansion databases such as MINES (Metabolic in silico Network Expansion Databases) can be searched (<http://minedatabase.mcs.anl.gov/>). MINES covers over 500,000 substances derived from KEGG and other pathway databases by applying known enzymatic transformation rules [40]. These novel compounds are not covered in traditional databases such as PubChem but can be utilized as hypothesized starting molecules for structure elucidation [45].

3. Mass Spectral Database Search for Fast Annotations

Mass spectral database search is currently the fastest and most accurate way for initial compound annotations. Current public and commercial mass spectral databases contain around 1–2 million spectra of one million unique compounds. Most of these spectra are EI mass spectra for GC-MS, while fewer are available for LC-MS/MS analysis. Traditionally, these databases have been derived from authentic experimental reference compounds and were collected from the literature [46]. Lately, computationally generated in silico spectra have also gained in importance, as discussed below. The experimentally derived as well as the in silico generated databases are enriched with metadata such as instrument types, collision energies, ionization mode and structural information such as the InChIKey [47] and SPLASH (spectral hash code) for uniqueness calculations [48]. Both InChIKey and SPLASH are important as unique identifiers in the structural and spectral domain. Errors during reference library building can be curated using software or manual data correction [49]. Table 3 lists

a selection of commonly used mass spectral databases, see recent reviews for a complete coverage of mass spectral databases [19,50].

Table 3. Overview of selected mass spectral databases commonly used for compound annotations. Specialized reviews that cover other mass spectral databases are referenced in the text.

Database	Targets	Description
NIST	EI-MS, CID-MS/MS	Curated DB, graphical interface
WILEY	EI-MS, CID-MS/MS	Largest collection of EI-MS data
METLIN [51]	CID-MS/MS	Developed for QTOF instruments
MoNA	EI, MS/MS, MSn	Autocurated collection of spectra
MassBank [52]	EI, MS/MS, MSn	Longest standing community database
mzCloud [53]	MSn	Multiple stage MSn
GNPS [54]	MS/MS	Community database
ReSpect [55]	MS/MS, RT	Plant metabolomics database

In terms of coverage, up to 400 metabolites were identified from NIST plasma reference standards utilizing multiple platforms and database matching [56]. The NIH Common Fund metabolomics ring trial with the participation of multiple US labs annotated around 1000 metabolites using multiple technologies and reference spectra matching. However, literature references for the plasma or serum metabolome covered up to 5000 compounds by combining targeted and non-targeted metabolomics analysis from five platforms [57]. It is therefore clear that matching experimental reference spectra to experimental reference databases is a severely limited process and covers only a fraction of the detectable metabolome.

Many modern algorithms for peak detection and mass spectral deconvolution have in-built database search algorithms. That includes freely available search algorithms such as the NIST MS Search GUI (graphical user interface), NIST MS PepSearch or MS-DIAL [58]. Commercial software from mass spectrometry vendors use similar algorithms.

Scoring mass spectra has been traditionally performed by a number of algorithms such as probability match searching, dot-product search and other similarity measures [19]. Recently, a novel hybrid similarity search method has been introduced that can annotate unknown spectra. The method does not account for the precursor m/z and instead utilizes similar neutral losses and fragmentation patterns [59]. Spectral similarity can also indicate structural similarity and this information can be used for annotation of unknown compounds [60]. Clustering approaches that use the cosine similarity of product ion spectra by clustering structurally similar compounds can improve the annotation of unknown metabolites [61]. Despite the advantages of a fast library search, it is becoming clear that mass spectral scoring algorithms have to be improved, especially for product ion spectra that contain only few fragments [46] or for those libraries that integrate spectra from multiple instrumentation types. Here, approaches that can calculate false discovery rates (FDR) will be useful to improve spectral match and annotation quality [62,63].

Community efforts have positively impacted the sharing of mass spectra. The MassBank database (<http://massbank.jp>) is one of the most successful examples, with a wide user base and contributors from many different countries [52]. In a coalition of database servers, the European MassBank efforts (<https://massbank.eu/>) [64] and MassBank of North America (<http://massbank.us/>) enable immediate sharing of mass spectra of annotated structures, including autocuration of spectra and chemical structure information (InChI keys). In comparison, the GNPS [54] spectral database utilizes crowd sourcing approaches to annotate unknown compounds. Commercial libraries such as NIST17 still play an important role because of high levels of manual curation, overall good data quality and wide coverage of substances.

4. In Silico Generation of Mass Spectra and MS/MS Spectra

As described before, scientists today have access to around 100 million known compounds in PubChem and ChemSpider. However, fewer than one million compounds have associated electron ionization (EI) mass spectra (for GC-MS applications), and even fewer LC-MS/MS tandem mass spectra are available. Generating in silico mass spectra, therefore, is a unique opportunity to close this gap. Research into computational generation of mass spectra has gained much traction during the last five years. Four general methods can be distinguished: quantum chemistry, machine learning, heuristic-based methods and chemical reaction-based methods.

Quantum chemistry methods use first-principles and purely physical and chemical information to generate mass spectra. In a major breakthrough for computational mass spectrometry, Grimme described in 2013 how Born–Oppenheimer ab initio molecular dynamics can be used to generate in silico electron ionization mass spectra of any given compound [65–69]. An overview of methods for in silico generation of mass spectra, including commercially or freely available algorithms is listed in Table 4.

Table 4. Overview of methods for in silico generation of mass spectra, including commercially or freely available algorithms. Additional tools are referenced in text.

In Silico Method	Software	Platform	Description
Quantum chemistry	QCEIMS	EI-MS	Uses chemistry first principles; requires cluster computations
Machine learning	CFM-ID/ CSI:FingerID	EI-MS CID-MS/MS	Requires diverse training sets; Fast method
Heuristic approaches	LipidBlast	CID-MS/MS	for specific compound classes (lipids); Fast method
Reaction chemistry methods	MassFrontier	EI-MS CID-MS/MS	generates only bar code spectra; Covers experimental gas phase reactions

Machine learning-based methods such as CFM-ID developed by Allen et al. allow for the computation of CID-MS/MS [70] and EI-MS spectra [71] directly from molecular structures. It is a very versatile approach useful for small molecules and peptides up to 1000 Da [72]. The methodology requires diverse and large training sets which subsequently will improve overall accuracy during training.

Heuristic approaches such as LipidBlast are advantageous for compound classes that have reoccurring and predictive fragmentations such as lipids [73]. However, the heuristic approach cannot be expanded to molecules with very diverse structural scaffolds. The libraries themselves can be easily extended to include novel or recently discovered lipid classes [40,45,74].

Reaction-based approaches are covered in the Mass Frontier software (HighChem Ltd., Bratislava, Slovakia) and based on thousands of reactions discovered in the literature. Novel molecules can be fragmented based on observed reaction pathways. Only bar code spectra can be generated, hence peak abundances are missing.

The accuracy of in silico generated peaks and their abundances have to be largely improved. A comparison between QCEIMS and CFM-ID has shown that both algorithms perform well enough to get correct identifications for half of the 61 investigated molecules [75]. However, certain rearrangement reactions, including McLafferty rearrangements, remain underestimated. The highly accurate and fast OM2 and OM3 semiempirical methods [76] have been further improved by the GFN-xTB Hamiltonian into QCEIMS [77]. Independent approaches described DFT reaction pathway and transition state modelling to model EI mass spectra [78] or Monte Carlo sampling to obtain EI mass spectra for select cases [79].

Currently, there is no fully automatic software for the generation of *in silico* MS/MS spectra based on LC-MS collision induced dissociation (CID). Several groups have shown interest in this challenging topic and have provided steps that can finally lead to a fully automated stand-alone solution. That includes workflows to automatically find the correct protonation sites in a molecule [80,81], ways to utilize rotamers, conformers, Boltzmann averaging and the evaluation of semiempirical and density functional methods (DFT) to calculate fragments.

The validation of generated *in silico* spectra is probably the most crucial aspect, especially when ‘blindly’ applying software models to large molecule repositories. For example, the original CFM-ID models were trained on a number of small metabolites. Therefore, these initial models are focused on lower molecular weight molecules and may not be feasible for the generation of *in silico* spectra of high molecular weight lipids or large complex secondary metabolites. In order to obtain high accuracy, the CFM-ID models have to be retrained with adequate lipid and secondary metabolite training sets. As always, external validation with mass spectra that were not available during training is highly recommended. For *ab initio* models, large validation sets with thousands of compounds have to be generated to obtain confidence scores.

Furthermore, regarding *in silico* spectra, two major problems will arise in the future. First, calculational processes follow the normal distribution; hence a large number of average accuracy *in silico* spectra will be observed. The flanks will consist of a small number of inaccurate spectra as well as a small number of high-quality spectra. Here, research needs to focus on ways to improve the average accuracy of *in silico* spectra predictions, but also to exclude such low-quality *in silico* spectra. In addition, the community will need to develop improved MS/MS match confidence scores. Otherwise, wrong spectra and publications with false compound annotations lead to many false-positive annotations in databases. The second problem is the generation of millions of very similar *in silico* spectra, because compound databases host millions of structurally very similar compounds. This will lead to an effect called database poisoning, filling mass spectral databases with compound spectra that cannot be easily distinguished by database search alone. Here, research has to focus on orthogonal filtering methods such as ion mobility or retention time filters.

5. In Silico Fragmentation Software

In silico fragmentation approaches for the annotation of unknown molecules are used in those cases where no reference mass spectra are available for database matching [82]. These generally involve matching experimental spectra against a selection of *in silico* generated fragments calculated on candidates retrieved from known compound databases (see Figure 2). Instead of searching mass spectral databases which cover only one million compounds, *in silico* fragmentation algorithms have access to molecular structure databases including ChemSpider and PubChem covering almost 100 million compounds [83].

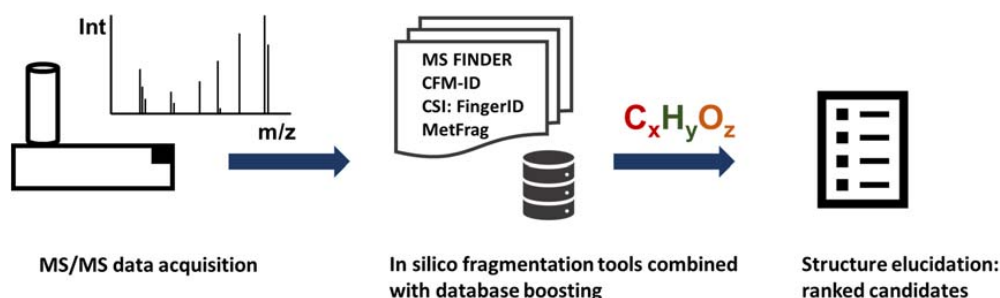


Figure 2. *In silico* fragmentation tools such as MS-Finder, CFM-ID, CSI:FingerID and Metfrag utilized known compounds from structure databases to calculate fragments compare those theoretical fragmentations against experimental spectra. When combined with MS/MS database search and utilizing additional metadata annotation rates can be increased tremendously.

These *in silico* fragmentation approaches aim to identify “known unknowns”—i.e., compounds present in molecular structure databases but without any reference spectra—by calculating a score between the experimental spectra and the predicted spectra (or predicted fragments). The major disadvantage is that “unknown–unknown” compounds cannot be elucidated in such a way. Below, we discuss some of the tools that have participated in structure elucidation challenges and can be used for batch annotations of unknown compounds (see Table 5). Additional software including iMet [84], MAGMa [85], MIDAS [86] and Midas-G [87] are discussed elsewhere. Most of the approaches below have been discussed in much greater technical detail in a series of excellent reviews [88–91].

Table 5. Selection of *in silico* fragmentation software, including commercially or freely available algorithms. Additional algorithms are referenced in the text.

Tools	Fragmentation Method	Compound DB	Type of Interface
MS-FINDER	Rule-based (hydrogen rearrangement rules)	15 integrated target DBs plus MINE and PubChem	Windows GUI
CFM-ID	Hybrid rule-based machine learning	KEGG, HMDB	Web application and command line tool
MetFrag	Hybrid rule-based combinatorial	HMDB, KEGG, PubChem	Web application, command line tool,
Mass Frontier	Rule-based (literature reaction mechanisms)	Internal MS database	Windows GUI
ChemDistiller	Fingerprint and spectral machine learning	17 different target databases, 130 Mio compounds total	Command line, web-based output
MAGMa, MAGMa+	Rule-based	PubChem, KEGG, HMDB	Web application, command line tool
CSI:FingerID	Combination of fragmentation trees and machine learning	PubChem and multiple bio databases	Platform independent GUI, command line tool

MetFrag [92] is a combinatorial fragmenter that retrieves candidate structures from PubChem, ChemSpider, KEGG, and a few other more specific compound databases. Candidates are fragmented using a bond dissociation approach and are finally matched to experimentally obtained spectra. MetFrag and MetFusion [93] have been actively developed and improved, allowing local or web-based use. The LipidFrag tool was developed later to increase confidence in lipid annotations [94].

MS-FINDER [84] is a Windows based GUI software aiding the structure elucidation process by *in silico* fragmentation of all predicted molecular formulas, determined from the accurate mass, isotope ratio, and product ion information [95], which are retrieved from 15 databases that are embedded into MS-FINDER [96,97]. The structures are then ranked by variety of factors including nine hydrogen rearrangement rules as the most contributing factor to the final score calculations.

CSI:FingerID [98] is a freely available web-service and uses a two-step scheme: first, a kernel-based approach is utilized to predict molecular fingerprints [99] from its MS/MS spectrum and then the predicted molecular fingerprints are matched against a molecular compound database. Included is a module that combines computation and comparison of fragmentation trees for the prediction of molecular properties of the unknowns as well as the molecular formula generation. Novel algorithms such as IOKR (input output kernel regression) [100] are now integrated into the workflow. The stand-alone SIRIUS GUI software [101] is used to calculate fragmentation trees and, subsequently, molecular formulas [102]. SIRIUS is now directly coupled to the CSI:FingerID online server that matches fingerprints against a database and retrieves ranked structure candidates.

CFM-ID (competitive fragmentation modeling) is a suite of software tools that can perform spectra prediction and compound identification. It is based on a machine-learning approach including chemical rules and is available for ESI MS/MS data as well as EI mass spectra. CFM-ID can be used as a web server or can be called locally through command line utilities on Windows, Linux and MacOS. For larger datasets, the software can be deployed to clusters to reduce the computational times.

ChemDistiller [103] is a Python-based tool that uses structural fingerprints and fragmentation patterns together with a machine learning algorithm to annotate unknown compounds. It utilizes multiple target databases covering more than 130 million compounds to annotate unknowns and the output is presented in a web interface for further inspection. It is a very fast and highly parallelized tool that makes use of modern multi-core CPUs.

Mass Frontier [82], developed by HighChem, is based on observed experimental gas-phase fragmentation reactions. It contains basic fragmentation rules as well as an exhaustive library of over 100,000 known fragmentation rules collected from published data which also allows for fragmentation predictions and annotation of unknowns [104]. The software supports electron ionization (EI) and collision induced dissociation (CID) ESI MS/MS modes. Mass Frontier can search internal databases or the mzCloud database and is commercially available.

To improve the annotation rates, database type restrictions such as environmental, plant, metabolic pathway databases can be applied. Taxonomy restrictions are also useful when researching specific organisms. Generally, *in silico* fragmentation algorithms still need to improve tremendously. A comparison of four algorithms using the CASMI test compounds as input has shown that pure *in silico* algorithms could only identify 17–25% of the compounds correctly [105]. Boosting the output by adding MS/MS search and bio-database focused lookups as well as combining the outputs of multiple software tools led to much higher identification rates of up to 93% accuracy [106]. Combining multiple *in silico* fragmentation software with *a-priori* information is a valuable option when facing a structure elucidation challenge [106].

6. Retention Time Prediction

Retention times are important as orthogonal filters during the structural determination in metabolic profiling experiments. A number of MS/MS and retention time databases have been developed for metabolic profiling [55]. However, these tools usually contain only a few hundred experimentally obtained retention time values. It is therefore useful to predict theoretical retention times utilizing the millions of existing compounds in compound databases by quantitative structure-retention relationship (QSRR) modelling [107]. This field of research has been active for more than 30 years. Traditionally, group-contribution methods were used for GC-MS modelling by assigning small retention index increments to specific substructures [108]. However, a vast amount of different separation columns and an infinite combination of solvent buffer systems and chromatographic conditions exist in LC-MS, locking the predicted models to very specific conditions [109].

Another major reason why there is no universal retention prediction method for LC-MS/MS is the lack of large and diverse training sets. A minimum of a thousand compounds covering all major chemical scaffolds in hydrophilic interaction liquid chromatography (HILIC) or reversed-phase chromatography (RP) are required to generate a robust retention prediction model useful for metabolic profiling.

An additional important consideration for retention time models is the applicability domain or structural space used in model building [110]. In short, if a natural product training set is used, it should be used for the prediction of natural product predictions and not for drugs. A simple measure would be to perform a principal component analysis on the substructure feature space for training samples and new predictive compounds and to confirm that the space overlaps. However, a recent approach utilized 1955 synthetic screening compounds that cover a similar scaffold space as small metabolites and used artificial neural networks to predict LC-MS retention indices for 202 endogenous metabolites [111]. This approach is particularly interesting because plated screening compounds are commonly less expensive than endogenous metabolites. By massively increasing the structural scaffold space, the retention model can become more robust, even if these molecules will never be annotated in biological samples. Many retention time prediction models are usually locked to a specific LC column and a solvent and buffer system, unless a “retention projection” method can be applied to transfer data to other chromatographic systems [112–114].

Retention times can be predicted by using chemical descriptors as input parameters which can be computed directly from structures by tools such as Dragon [115], MOLD2 [116] or PaDel [117]. Dragon 7 now calculates 5270 molecular descriptors, covering fragment counts, topological and geometrical descriptors. Low-energy three dimensional conformer structures can be generated by a number of tools [118] and even better with quantum chemical methods [119]. Subsequently, regression models can be built using the descriptor data as input and the retention time as a target function. Over 200 machine learning models, preferably with deep neural networks [120] or fast random forest methods [121], are now available. To improve accuracy and prediction power, complex gradient boosting methods (XGBoost/LightGBM) and ensemble methods such as bagging, stacking and averaging are now routinely employed [122]. In the past, a wide variety of retention prediction models have been proposed for HILIC and reversed phase columns based on different machine learning approaches. These included partial least square methods [123–125], multiple linear regression [126–128], support vector regression [129,130], random forests [131] and artificial neural networks [132–134].

In summary, the success of the retention time modelling depends on the size and the diversity of the compound training data set. Currently, most RT models are locked to specific columns and conditions, unless a retention projection method is used. For useful retention time prediction models, the only remedies are large and diverse training sets covering multiple compound classes to obtain reliable, highly predictive and accurate models.

7. Ion Mobility and the Use of Collision Cross Section (CCS) Values

LC-MS/MS alone will often be unable to discriminate between stereoisomers and regioisomers, unless chiral columns are utilized. It is therefore useful to couple ion mobility analyzers to LC-MS/MS to allow for a higher number of features to be separated and detected [135]. Ion mobility is a technique that separates ions in an inert buffer gas (nitrogen, hydrogen) under the influence of an electric field [136,137]. Several types of ion mobility analyzers are available, among them drift tube ion mobility (DTIMS), traveling wave ion mobility spectrometry (TWIMS) and FAIMS [138].

For DTIMS and TWIMS, the observed drift times are influenced by relative molecule size and conformational parameters. For DTIMS, cross-section values (CCS) can be directly measured and computed [139,140], and for TWIMS the CSS values can be obtained from calibrations with known standards [141]. The FAIMS technology has limited peak capacity [142,143], but can be used as an orthogonal filter to separate different classes of compounds and to improve signal/noise ratios during measurements [144]. For FAIMS, no collision cross-section values (CCS) can be determined [138].

The experimental CCS values have a very high reproducibility and CCS values with relative standard deviation (RSD) of <1–2% can be routinely obtained [139,145,146]. This opens up the LC-IMS-MS/MS technology for orthogonal filtering approaches utilizing CSS values [147] (see Figure 3) and more importantly for predictive technologies utilizing CCS values in a similar to retention time predictions. Such predictive approaches can include computational and quantum chemical models [148,149] as well as machine learning predictions [150] such as artificial neural networks [132,151]. Prediction errors as low as 3% have been reported for CCS models [152]. Once these models are applied to structures from large metabolomic databases, they can be used for filtering during the compound identification process [138,153], and such predicted values are covered in publicly available databases such as MetCCS [152] or LipidCCS [154,155]. Currently, an estimated total of 3000–4000 experimental small molecule CCS values have been reported in a recent review [150] with the largest single collection containing CCS values for 1420 compounds [145]. Focused collections for sterols [156], metabolites and xenobiotics are also available [139,157].

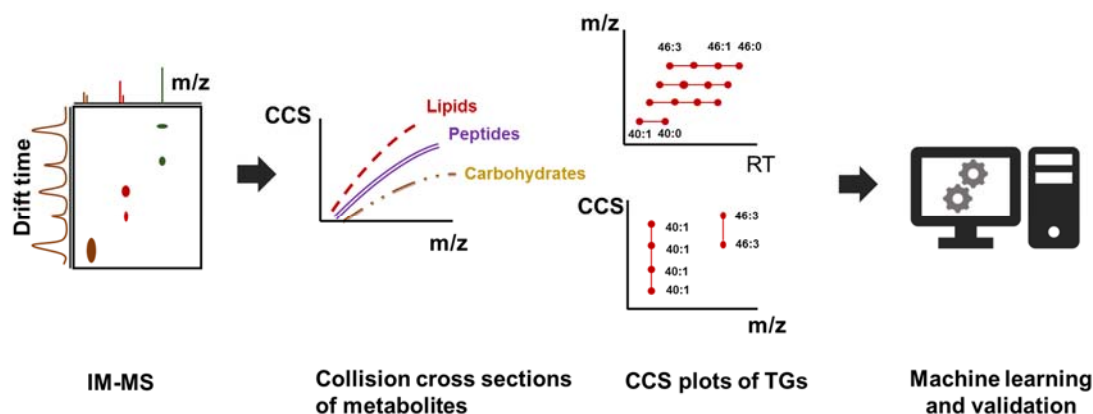


Figure 3. Ion mobility can be used as an additional orthogonal approach to resolve complex mixtures. The experimental collision cross-section values (CCS) can be further utilized to train machine learning models to further enrich compound databases with CCS information.

Several considerations have to be taken into account when working with CCS values and predictive databases. CCS values of individual compounds depend on many additional parameters such as buffer gas, solvents, temperature, pH, ion activation voltage and conformer/rotamer ensembles [158,159]. For example, different ion species such as $[M + H]^+$ and $[M + Na]^+$ have different CCS values, differing on average $\pm 7 \text{ \AA}^2$ based on values obtained from [139]. This is related to conformational changes and subsequently leads to the conclusion that different adducts have to be modelled and predicted separately. Furthermore, different protonation sites or protomers can lead to different CCS values [145]. Drugs such as benzocaine can have N- or O-protonated forms leading to different CCS values for the same compound [160]. The different protomers can be determined with the help of quantum chemical methods [161,162] and cheminformatics methods that calculate different protonation sites. Reference standards themselves may not be enantiomerically pure and therefore can lead to measurement of multiple experimental CCS values. Furthermore, while CCS values predicted on the same instrument type have low RSD measurement errors $<1\%$ [163], the experimental CCS values may differ between different instrumental setups (DTIMS/TWIMS), as well as prediction models. The drug Indomethacin for the proton adduct has a reported CCS value of 183.54 \AA^2 measured on a drift tube IMS (DTIMS) [139]; the same compound has a CCS value of 179.039 \AA^2 measured on a TWIMS setup, and the predicted value is 197.7 \AA^2 and therefore falls outside the 3% median prediction error [164].

Because of the IMS capability of separating stereoisomers and other isobaric compounds, the routine use of CCS values will become more and more important. The excellent experimental reproducibilities of CCS measurements compared to retention times will also improve identification rates. Once larger CCS datasets become publicly available, they can be combined, average consensus values can be calculated and CCS prediction methods can be retrained with larger compound numbers and therefore will automatically become more accurate. Technological advances such as printed circuit board (PCB)-based devices led to ion elevators and escalators in multilevel structures [165]. Therefore, such structures for lossless ion manipulations (SLIM) have demonstrated unprecedented ultra-high resolution ion mobility [166].

8. Compound Identification: Hybrid and Orthogonal Approaches

The following section discusses some general compound identification workflows as well as a few selected cases of single compound identification examples via mass spectrometry. Workflows are important for highly reproducible and repeatable metabolomics analysis. Among those are Galaxy workflows [161] such as Workflow4metabolomics.org, as well as Taverna and KNIME workflows, but with a considerably lower user base [167]. A conceptual compound ID workflow has been

described that includes in silico metabolic synthesis, in silico fragmentation [168] and finally annotation of compounds via database scoring [169]. The same paper discusses the importance of meta-integration of multiple tools and multiple layers of information to improve confidence in compound identification. Another related review discusses the importance of inclusion of MS¹ peak relationships such as adducts and neutral losses, the inclusion of MS/MS data and biochemical knowledge as well as modelling of retention times as an orthogonal filter. A knowledge-based workflow for metabolite annotations that includes ionization rules, adduct formation rules and retention time rules was described in [170].

However, even in-source fragmentation LC-MS mass spectra when used together with retention times of authentic compounds can be sufficient for 'Level 1' annotations in metabolomics [171]. A pipeline that uses multicriteria scoring, including retention times, intensity profiles and adduct patterns was developed for high-resolution mass spectral data [172]. The extraction of common occurring substructures from MS/MS data can help during higher level annotations [173]. Another workflow included the use of multiple identification criteria such as accurate mass, retention time, MS/MS spectrum, and product/precursor ion intensity ratios to support reversed phase and HILIC based metabolic profiling [174]. Two in silico fragmenters and two retention prediction models were utilized to annotate hydrophobic compounds [175]. A tool for improved and automated adduct detection was discussed in [176], leading to 83% correct annotations of adduct ions. The dereplication of natural products with the help of a fragment database was described in [177]. Pitfalls, limitations and general recommendation during data processing and compound identifications were discussed in [24,178,179].

Full structure elucidation of single novel compounds with chromatography and mass spectrometric analysis is possible but is harder than with the isolation of compounds and NMR analysis. A clear benefit of LC-MS/MS approaches is the limited amount of material needed, in comparison to LC-MS/MS-NMR methods. A recent report annotated N¹-acetylisoptreanine and N¹-acetylisoptreanine-gamma-lactam by metabolic profiling and used custom synthesis to confirm the commercially unavailable metabolite [180]. Another approach used multiple-stage tandem mass spectrometry (MS⁴) and custom synthesis to identify and confirm *N,N,N*-trimethyl-L-alanyl-L-proline betaine in human plasma. Novel glycolipids were found in yeast annotated by combining multiple mass spectrometric platforms and chiral chromatography to ascertain stereoisomer configuration [181]. Another approach showed the combined use of high-resolution MS/MS data and use of the metabolic in-silico network expansion database (MINE) for the discovery of novel methylated epi-metabolites including N-methyl-UMP [45]. Natural products can be manually annotated with high success rates [182], but such approaches require deep mass spectral knowledge. In the future, such manual approaches must be translated into practical expert-algorithms and software that allows non-experts to perform such complicated analysis to a certain degree [27]. Finally, all pipelines and workflows must be validated by independent and external benchmark sets such as the CASMI competitions discussed below.

9. Critical Assessment of Small Molecule Analysis (CASMI)

The CASMI (critical assessment of small molecule identification) contest (<http://www.casmi-contest.org>) has been held since 2012 as a worldwide scientific competition to determine the best approaches for identifying small molecule structures directly from mass spectra [183,184]. The competitions are commonly structured into different categories, including best natural product determination [96,182,185], best molecular formula determination [186] and unknown compound determination. More recently, categories that allow for in silico fragmentation software only [187] and a category that allows for all meta-data use were included [85]. Participants publish their findings in special journal issues selected by the CASMI organizers and describe how they implemented and performed their structure annotation processes.

The latest CASMI 2017 contest featured 300 small molecule challenges and may continue to serve as a test bed for the performance and comparison of software tools and pipelines. On the other hand,

many research papers describe approaches and pipelines that focus on a few selected “cherry picked” test cases. Therefore, it is recommended for groups that develop compound identification software to participate in the yearly CASMI contests to showcase the performance of their software against others. Best of all, any published article about novel approaches or software tools should participate in the CASMI small molecule identification contests or at least use former CASMI data sets for validation of the approaches used.

Future CASMI contests may be held in a completely automatic fashion, as long as the software and pipeline are fully publicly available. One idea would be to make these tools so easy to use that non-specialists from the broader community can utilize them quickly and improve compound identification rates. The increasing number of challenges and CASMI participants shows that the field of unknown-identification is moving steadily forward.

10. Data Sharing and Data Retention

Sharing research data and software helps to validate the claims made in publications and, more importantly, lets researchers freely reuse that data and develop novel research ideas [188]. Unfortunately, while journals support data sharing, they often do not strictly enforce it [189]. Here, funding agencies such as the National Institutes of Health (NIH) in the United States have a large leverage to make data sharing mandatory. Both NIH and the US National Science Foundation (NSF) require data retention and data sharing plans for grant proposals, cultivating a way for better reuse of research data. Currently, funding organizations worldwide do not strictly enforce the public sharing of metabolomics data. This is contrary to genomics, where deposition of genomic data is required before any publication.

For computational tools and software, it is recommended to use public software repositories such as GitHub, BitBucket and SourceForge services (see Table 6). In this case, repositories can be forked (copied) and multiple copies remain even when the original distributor does not support them anymore.

Table 6. Overview of collaborative software and data sharing repositories, major metabolomics repositories and mass spectral sharing initiatives.

Data Sharing	Link	Description
GitHub	github.com	Software development platform
BitBucket	bitbucket.org	Collaborative software sharing
SourceForge	sourceforge.net	Collaborative software sharing
Zenodo	zenodo.org	Open research data repository
Figshare	figshare.com	Online research data repository
Metabolomics Workbench	metabolomicsworkbench.org	Experimental metabolomics data
MetaboLights	ebi.ac.uk/metabolights	European metabolomics repository
OpenMSI	openmsi.nersc.gov	Mass spectral imaging data
MetaSpace	metaspace2020.eu	Mass spectral imaging data
GNPS	gnps.ucsd.edu	Mass spectral data sharing
MassBank	massbank.jp	Mass spectral data sharing
MoNA	massbank.us	Mass spectral sharing community
Norman MassBank	massbank.eu	Mass spectral data sharing

For metabolomics data sets, the Metabolomics Workbench [30] or the European metabolomics repository MetaboLights [31] should be considered. These repositories contain a high level of metadata information, which requires a high level of data preparation before the upload process. The advantage is that experiments are very well described and that such metadata can be queried at a later time point. The incentive of the GNPS repository [54] is that mass spectra of many unknown compounds are collected, and identification of such spectra might be enhanced through community efforts. The OpenMSI [190] and Metaspace.eu [63] projects provide open analysis solutions for mass spectral

imaging data. Scientific data sets from all branches of research can be submitted to the Zenodo research repository, which also supports citable digital object identifiers (DOI). The long-standing effort of collecting freely available mass spectra of pure reference compounds at MassBank (Japan) has now been complemented by collaborative efforts in the USA (MassBank of North America, MoNA) and the Norman MassBank in Europe. Due to the allowed unrestricted use, open spectral collections can be used for algorithm training in open or commercial software.

Specifically, the MoNA database has an advantage of automated spectral uploads via REST API, which allows for instantaneous sharing of novel compounds and associated spectra. MoNA collates all worldwide publicly available mass spectra, including spectra from MetaboBASE, GNPS, HMDB, LipidBlast, ReSpec and MassBank spectra in one unique repository. Users can freely download spectra based on metadata tags, including based on instrument, vendor, mass accuracy, types of chromatography, or based on compound classes (supported by ClassyFire) [191].

The publication of tools or databases that are neither publicly nor commercially available should be avoided. Such opaque software does not contribute much to the field and cannot be validated independently. We therefore mostly refrained from referencing such publications or tools in this review. Software tools should be validated on public, large and diverse datasets before making claims that they outperform any other tool.

11. Conclusions and Outlook

Computational metabolomics strategies for compound identification have gained increased attention in the community. Unknown metabolite signals cannot easily be used for biological interpretations [7], and increased efforts and validations for compound identifications are critical for the field to move forward. Approaches that do not require the identification of metabolic features should be used with extreme caution because they may lead to false interpretations. The identification of metabolites with a high level of confidence is required in order to improve metabolomics applications in the field of translational and clinical research.

Bioinformatics researchers have helped the proteomics and genomics community over many years to solve problems in their domain. However, the bioinformatics community had a smaller impact on the small-molecule community due to the chemical structure-centric approaches that are needed for structure elucidation in metabolomics. To this end, the much smaller cheminformatics community still struggles to provide adequate support simply due to its much smaller size and impact. Therefore, collaboration with researchers from scientific branches such as machine learning and the quantum chemistry community need to be actively embraced. The computational metabolomics community is a quite small but innovative community, and many more research groups worldwide contribute now in friendly competition.

Author Contributions: I.B., T.K., J.J. and O.F. wrote the paper in a collaborative approach. All authors read and approved the final version of the manuscript.

Acknowledgments: Funding was provided by the US National Science Foundation projects MCB 113944 and MCB 1611846 to O.F. and the US National Institutes of Health U24 DK097154 to O.F. Additional funding for T.K. was provided by the American Heart Association grant 15SDG25760020 (Irvin) and NIH 7R01HL091357-06 (Arnett). We are thankful to Boris Šlogar for revision and linguistic editing efforts.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations and Glossary

MS ⁿ	Multiple stage mass spectrometry
CASMI	Critical Assessment of Small Molecule Identification
CCS	Collisional cross-section
CFM-ID	Competitive Fragmentation Modeling for Metabolite Identification
FAHFAs	Fatty Acid ester of Hydroxyl Fatty Acids
Fragmentation tree	Mass spectral fragmentation pathway of a compound

GNPS	Global Natural Products Social molecular networking
HMDB	Human Metabolome Database
IM	Ion mobility
InChIKey	Hash key or short unique structure code
LipidBlast	In silico generated database for lipid identification
MassBank	Mass spectral database
MetaboBASE	Mass spectral library developed by Bruker
MoNA	MassBank of North America
NIST	National Institute of Standards and Technology
NMR	Nuclear Magnetic Resonance
ReSpect	RIKEN MSn spectral database for phytochemicals
SPLASH	Hashed code or unique identifier for mass spectra

References

- James, S.J.; Cutler, P.; Melnyk, S.; Jernigan, S.; Janak, L.; Gaylor, D.W.; Neubrander, J.A. Metabolic biomarkers of increased oxidative stress and impaired methylation capacity in children with autism. *Am. J. Clin. Nutr.* **2004**, *80*, 1611–1617. [[CrossRef](#)] [[PubMed](#)]
- Vasan, R.S. Biomarkers of cardiovascular disease: Molecular basis and practical considerations. *Circulation* **2006**, *113*, 2335–2362. [[CrossRef](#)] [[PubMed](#)]
- Beger, R.D.; Dunn, W.; Schmidt, M.A.; Gross, S.S.; Kirwan, J.A.; Cascante, M.; Brennan, L.; Wishart, D.S.; Oresic, M.; Hankemeier, T.; et al. Metabolomics enables precision medicine: “A White Paper, Community Perspective”. *Metabolomics* **2016**, *12*, 149. [[CrossRef](#)] [[PubMed](#)]
- Wishart, D.S. Computational strategies for metabolite identification in metabolomics. *Bioanalysis* **2009**, *1*, 1579–1596. [[CrossRef](#)] [[PubMed](#)]
- Da Silva, R.R.; Dorrestein, P.C.; Quinn, R.A. Illuminating the dark matter in metabolomics. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12549–12550. [[CrossRef](#)] [[PubMed](#)]
- Peisl, L.; Schymanski, E.L.; Wilmes, P. Dark matter in host-microbiome metabolomics: Tackling the unknowns—A review. *Anal. Chim. Acta* **2017**. [[CrossRef](#)]
- Uppal, K.; Walker, D.I.; Liu, K.; Li, S.; Go, Y.-M.; Jones, D.P. Computational metabolomics: A framework for the million metabolome. *Chem. Res. Toxicol.* **2016**, *29*, 1956–1975. [[CrossRef](#)] [[PubMed](#)]
- Kim, H.K.; Choi, Y.H.; Verpoorte, R. NMR-based metabolomic analysis of plants. *Nat. Protoc.* **2010**, *5*, 536–549. [[CrossRef](#)] [[PubMed](#)]
- Eisenreich, W.; Bacher, A. Advances of high-resolution NMR techniques in the structural and metabolic analysis of plant biochemistry. *Phytochemistry* **2007**, *68*, 2799–2815. [[CrossRef](#)] [[PubMed](#)]
- Pérez-Victoria, I.; Martín, J.; Reyes, F. Combined LC/UV/MS and NMR strategies for the dereplication of marine natural products. *Planta Med.* **2016**, *82*, 857–871. [[CrossRef](#)] [[PubMed](#)]
- Hubert, J.; Nuzillard, J.-M.; Renault, J.-H. Dereplication strategies in natural product research: How many tools and methodologies behind the same concept? *Phytochem. Rev.* **2017**, *16*, 55–95. [[CrossRef](#)]
- Schymanski, E.L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H.P.; Hollender, J. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environ. Sci. Technol.* **2014**, *48*, 2097–2098. [[CrossRef](#)] [[PubMed](#)]
- Rochat, B. Proposed Confidence Scale and ID Score in the Identification of Known-Unknown Compounds Using High Resolution MS Data. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 709–723. [[CrossRef](#)] [[PubMed](#)]
- Sumner, L.W.; Amberg, A.; Barrett, D.; Beale, M.H.; Beger, R.; Daykin, C.A.; Fan, T.W.-M.; Fiehn, O.; Goodacre, R.; Griffin, J.L. Proposed minimum reporting standards for chemical analysis. *Metabolomics* **2007**, *3*, 211–221. [[CrossRef](#)] [[PubMed](#)]
- Viant, M.R.; Kurland, I.J.; Jones, M.R.; Dunn, W.B. How close are we to complete annotation of metabolomes? *Curr. Opin. Chem. Biol.* **2017**, *36*, 64–69. [[CrossRef](#)] [[PubMed](#)]
- Milman, B.L.; Zhurkovich, I.K. The chemical space for non-target analysis. *TrAC Trends Anal. Chem.* **2017**, *97*, 179–187. [[CrossRef](#)]
- Spicer, R.; Salek, R.M.; Moreno, P.; Canueto, D.; Steinbeck, C. Navigating freely-available software tools for metabolomics analysis. *Metabolomics* **2017**, *13*, 106. [[CrossRef](#)] [[PubMed](#)]

18. Cajka, T.; Fiehn, O. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Anal. Chem.* **2016**, *88*, 524–545. [[CrossRef](#)] [[PubMed](#)]
19. Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z.; Mehta, S.S.; Wohlgemuth, G.; Barupal, D.K.; Showalter, M.R.; Arita, M.; et al. Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom. Rev.* **2017**. [[CrossRef](#)] [[PubMed](#)]
20. Domingo-Almenara, X.; Montenegro-Burke, J.R.; Benton, H.P.; Siuzdak, G. Annotation: A Computational Solution for Streamlining Metabolomics Analysis. *Anal. Chem.* **2017**, *90*, 480–489. [[CrossRef](#)] [[PubMed](#)]
21. Kind, T.; Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal. Rev.* **2010**, *2*, 23–60. [[CrossRef](#)] [[PubMed](#)]
22. Vaniya, A.; Fiehn, O. Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics. *Trends Anal. Chem.* **2015**, *69*, 52–61. [[CrossRef](#)] [[PubMed](#)]
23. Misra, B.B. New tools and resources in metabolomics: 2016–2017. *Electrophoresis* **2018**. [[CrossRef](#)] [[PubMed](#)]
24. Fenaille, F.; Saint-Hilaire, P.B.; Rousseau, K.; Junot, C. Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand? *J. Chromatogr. A* **2017**, *1526*, 1–12. [[CrossRef](#)] [[PubMed](#)]
25. Zaikin, V.; Halket, J.M. *A Handbook of Derivatives for Mass Spectrometry*; IM Publications: West Sussex, UK, 2009.
26. Gil de la Fuente, A.; Armitage, E.G.; Otero, A.; Barbas, C.; Godzien, J. Differentiating signals to make biological sense—A guide through databases for MS-based non-targeted metabolomics. *Electrophoresis* **2017**. [[CrossRef](#)] [[PubMed](#)]
27. Aksenov, A.A.; da Silva, R.; Knight, R.; Lopes, N.P.; Dorrestein, P.C. Global chemical analysis of biology by mass spectrometry. *Nat. Rev. Chem.* **2017**, *1*, 0054. [[CrossRef](#)]
28. Garg, N.; Luzzatto-Knaan, T.; Melnik, A.V.; Caraballo-Rodríguez, A.M.; Floros, D.J.; Petras, D.; Gregor, R.; Dorrestein, P.C.; Phelan, V.V. Natural products as mediators of disease. *Nat. Prod. Rep.* **2017**, *34*, 194–219. [[CrossRef](#)] [[PubMed](#)]
29. Bloszies, C.S.; Fiehn, O. Using untargeted metabolomics for detecting exposome compounds. *Curr. Opin. Toxicol.* **2018**, *8*, 87–92. [[CrossRef](#)]
30. Sud, M.; Fahy, E.; Cotter, D.; Azam, K.; Vadivelu, I.; Burant, C.; Edison, A.; Fiehn, O.; Higashi, R.; Nair, K.S. Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* **2015**, *44*, D463–D470. [[CrossRef](#)] [[PubMed](#)]
31. Haug, K.; Salek, R.M.; Conesa, P.; Hastings, J.; de Matos, P.; Rijnbeek, M.; Mahendrakar, T.; Williams, M.; Neumann, S.; Rocca-Serra, P. MetaboLights—An open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* **2012**, *41*, D781–D786. [[CrossRef](#)] [[PubMed](#)]
32. Bolton, E.E.; Wang, Y.; Thiessen, P.A.; Bryant, S.H. PubChem: Integrated platform of small molecules and biological activities. *Annu. Rep. Comput. Chem.* **2008**, *4*, 217–241.
33. Pence, H.E.; Williams, A. ChemSpider: An online chemical information resource. *J. Chem. Educ.* **2010**, *87*, 1123–1124. [[CrossRef](#)]
34. Kanehisa, M.; Goto, S.; Hattori, M.; Aoki-Kinoshita, K.F.; Itoh, M.; Kawashima, S.; Katayama, T.; Araki, M.; Hirakawa, M. From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res.* **2006**, *34*, D354–D357. [[CrossRef](#)] [[PubMed](#)]
35. Caspi, R.; Foerster, H.; Fulcher, C.A.; Kaipa, P.; Krummenacker, M.; Latendresse, M.; Paley, S.; Rhee, S.Y.; Shearer, A.G.; Tissier, C. The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **2007**, *36*, D623–D631. [[CrossRef](#)] [[PubMed](#)]
36. Schomburg, I.; Jeske, L.; Ulbrich, M.; Placzek, S.; Chang, A.; Schomburg, D. The BRENDA enzyme information system—From a database to an expert system. *J. Biotechnol.* **2017**, *261*, 194–206. [[CrossRef](#)] [[PubMed](#)]
37. Wishart, D.S.; Jewison, T.; Guo, A.C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E. HMDB 3.0—The human metabolome database in 2013. *Nucleic Acids Res.* **2012**, *41*, D801–D807. [[CrossRef](#)] [[PubMed](#)]
38. Degtyarenko, K.; De Matos, P.; Ennis, M.; Hastings, J.; Zbinden, M.; McNaught, A.; Alcántara, R.; Darsow, M.; Guedj, M.; Ashburner, M. ChEBI: A database and ontology for chemical entities of biological interest. *Nucleic Acids Res.* **2007**, *36*, D344–D350. [[CrossRef](#)] [[PubMed](#)]

39. Gu, J.; Gui, Y.; Chen, L.; Yuan, G.; Lu, H.-Z.; Xu, X. Use of natural products as chemical library for drug discovery and network pharmacology. *PLoS ONE* **2013**, *8*, e62839. [[CrossRef](#)] [[PubMed](#)]
40. Jeffries, J.G.; Colastani, R.L.; Elbadawi-Sidhu, M.; Kind, T.; Niehaus, T.D.; Broadbelt, L.J.; Hanson, A.D.; Fiehn, O.; Tyo, K.E.; Henry, C.S. MINEs: Open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics. *J. Cheminform.* **2015**, *7*, 44. [[CrossRef](#)] [[PubMed](#)]
41. Pye, C.R.; Bertin, M.J.; Lokey, R.S.; Gerwick, W.H.; Linington, R.G. Retrospective analysis of natural products provides insights for future discovery trends. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 5601–5606. [[CrossRef](#)] [[PubMed](#)]
42. O'Hagan, S.; Kell, D.B. Analysing and navigating natural products space for generating small, diverse, but representative chemical libraries. *Biotechnol. J.* **2018**, *13*, 1700503. [[CrossRef](#)] [[PubMed](#)]
43. Warth, B.; Spangler, S.; Fang, M.; Johnson, C.H.; Forsberg, E.M.; Granados, A.; Martin, R.L.; Domingo-Almenara, X.; Huan, T.; Rinehart, D. Exposome-scale investigations guided by global metabolomics, pathway analysis, and cognitive computing. *Anal. Chem.* **2017**, *89*, 11505–11513. [[CrossRef](#)] [[PubMed](#)]
44. Williams, A.J.; Grulke, C.M.; Edwards, J.; McEachran, A.D.; Mansouri, K.; Baker, N.C.; Patlewicz, G.; Shah, I.; Wambaugh, J.F.; Judson, R.S. The CompTox Chemistry Dashboard: A community data resource for environmental chemistry. *J. Cheminform.* **2017**, *9*, 61. [[CrossRef](#)] [[PubMed](#)]
45. Lai, Z.; Tsugawa, H.; Wohlgemuth, G.; Mehta, S.; Mueller, M.; Zheng, Y.; Ogiwara, A.; Meissen, J.; Showalter, M.; Takeuchi, K. Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat. Methods* **2018**, *15*, 53. [[CrossRef](#)] [[PubMed](#)]
46. Stein, S. Mass spectral reference libraries: An ever-expanding resource for chemical identification. *Anal. Chem.* **2012**, *84*, 7274–7282. [[CrossRef](#)] [[PubMed](#)]
47. Heller, S.; McNaught, A.; Stein, S.; Tchekhovskoi, D.; Pletnev, I. InChI-the worldwide chemical structure identifier standard. *J. Cheminform.* **2013**, *5*, 7. [[CrossRef](#)] [[PubMed](#)]
48. Wohlgemuth, G.; Mehta, S.S.; Mejia, R.F.; Neumann, S.; Pedrosa, D.; Pluskal, T.; Schymanski, E.L.; Willighagen, E.L.; Wilson, M.; Wishart, D.S. SPLASH, a hashed identifier for mass spectra. *Nat. Biotechnol.* **2016**, *34*, 1099–1101. [[CrossRef](#)] [[PubMed](#)]
49. Wallace, W.E.; Ji, W.; Tchekhovskoi, D.V.; Phinney, K.W.; Stein, S.E. Mass spectral library quality assurance by inter-library comparison. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 733–738. [[CrossRef](#)] [[PubMed](#)]
50. Vinaixa, M.; Schymanski, E.L.; Neumann, S.; Navarro, M.; Salek, R.M.; Yanes, O. Mass spectral databases for LC/MS-and GC/MS-based metabolomics: State of the field and future prospects. *TrAC Trends Anal. Chem.* **2016**, *78*, 23–35. [[CrossRef](#)]
51. Guigas, C.; Montenegro-Burke, J.R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A.E. METLIN: A Technology Platform for Identifying Knowns and Unknowns. *Anal. Chem.* **2018**, *90*, 3156–3164. [[CrossRef](#)] [[PubMed](#)]
52. Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; et al. MassBank: A public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.* **2010**, *45*, 703–714. [[CrossRef](#)] [[PubMed](#)]
53. Wang, J.; Peake, D.A.; Mistrik, R.; Huang, Y. A platform to identify endogenous metabolites using a novel high performance Orbitrap MS and the mzCloud Library. *Blood* **2013**, *4*, 2–8.
54. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapon, C.A.; Luzzatto-Knaan, T. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837. [[CrossRef](#)] [[PubMed](#)]
55. Sawada, Y.; Nakabayashi, R.; Yamada, Y.; Suzuki, M.; Sato, M.; Sakata, A.; Akiyama, K.; Sakurai, T.; Matsuda, F.; Aoki, T. RIKEN tandem mass spectral database (ReSpect) for phytochemicals: A plant-specific MS/MS-based data resource and database. *Phytochemistry* **2012**, *82*, 38–45. [[CrossRef](#)] [[PubMed](#)]
56. Simon-Manso, Y.; Lowenthal, M.S.; Kilpatrick, L.E.; Sampson, M.L.; Telu, K.H.; Rudnick, P.A.; Mallard, W.G.; Bearden, D.W.; Schock, T.B.; Tchekhovskoi, D.V.; et al. Metabolite profiling of a NIST Standard Reference Material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. *Anal. Chem.* **2013**, *85*, 11725–11731. [[CrossRef](#)] [[PubMed](#)]
57. Psychogios, N.; Hau, D.D.; Peng, J.; Guo, A.C.; Mandal, R.; Bouatra, S.; Sinelnikov, I.; Krishnamurthy, R.; Eisner, R.; Gautam, B. The human serum metabolome. *PLoS ONE* **2011**, *6*, e16957. [[CrossRef](#)] [[PubMed](#)]

58. Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **2015**, *12*, 523. [[CrossRef](#)] [[PubMed](#)]
59. Moorthy, A.S.; Wallace, W.E.; Kearsley, A.J.; Tchekhovskoi, D.V.; Stein, S.E. Combining fragment-ion and neutral-loss matching during mass spectral library searching: A new general purpose algorithm applicable to illicit drug identification. *Anal. Chem.* **2017**, *89*, 13261–13268. [[CrossRef](#)] [[PubMed](#)]
60. Schollée, J.E.; Schymanski, E.L.; Stravs, M.A.; Gulde, R.; Thomaidis, N.S.; Hollender, J. Similarity of High-Resolution Tandem Mass Spectrometry Spectra of Structurally Related Micropollutants and Transformation Products. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 2692–2704. [[CrossRef](#)] [[PubMed](#)]
61. Depke, T.; Franke, R.; Brönstrup, M. Clustering of MS2 spectra using unsupervised methods to aid the identification of secondary metabolites from *Pseudomonas aeruginosa*. *J. Chromatogr. B* **2017**, *1071*, 19–28. [[CrossRef](#)] [[PubMed](#)]
62. Scheubert, K.; Hufsky, F.; Petras, D.; Wang, M.; Nothias, L.-F.; Dührkop, K.; Bandeira, N.; Dorrestein, P.C.; Böcker, S. Significance estimation for large scale metabolomics annotations by spectral matching. *Nat. Commun.* **2017**, *8*, 1494. [[CrossRef](#)] [[PubMed](#)]
63. Palmer, A.; Phapale, P.; Chernyavsky, I.; Lavigne, R.; Fay, D.; Tarasov, A.; Kovalev, V.; Fuchser, J.; Nikolenko, S.; Pineau, C. FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nat. Methods* **2017**, *14*, 57. [[CrossRef](#)] [[PubMed](#)]
64. Stravs, M.A.; Schymanski, E.L.; Singer, H.P.; Hollender, J. Automatic recalibration and processing of tandem mass spectra using formula annotation. *J. Mass Spectrom.* **2013**, *48*, 89–99. [[CrossRef](#)] [[PubMed](#)]
65. Grimme, S. Towards first principles calculation of electron impact mass spectra of molecules. *Angew. Chem. Int. Ed.* **2013**, *52*, 6306–6312. [[CrossRef](#)] [[PubMed](#)]
66. Bauer, C.A.; Grimme, S. How to compute electron ionization mass spectra from first principles. *J. Phys. Chem. A* **2016**, *120*, 3755–3766. [[CrossRef](#)] [[PubMed](#)]
67. Bauer, C.A.; Grimme, S. First principles calculation of electron ionization mass spectra for selected organic drug molecules. *Org. Biomol. Chem.* **2014**, *12*, 8737–8744. [[CrossRef](#)] [[PubMed](#)]
68. Ásgeirsson, V.; Bauer, C.A.; Grimme, S. Unimolecular decomposition pathways of negatively charged nitriles by ab initio molecular dynamics. *Phys. Chem. Chem. Phys.* **2016**, *18*, 31017–31026. [[CrossRef](#)] [[PubMed](#)]
69. Bauer, C.A.; Grimme, S. Elucidation of electron ionization induced fragmentations of adenine by semiempirical and density functional molecular dynamics. *J. Phys. Chem. A* **2014**, *118*, 11479–11484. [[CrossRef](#)] [[PubMed](#)]
70. Allen, F.; Greiner, R.; Wishart, D. Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics* **2015**, *11*, 98–110. [[CrossRef](#)]
71. Allen, F.; Pon, A.; Greiner, R.; Wishart, D. Computational prediction of electron ionization mass spectra to assist in GC/MS compound identification. *Anal. Chem.* **2016**, *88*, 7689–7697. [[CrossRef](#)] [[PubMed](#)]
72. Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. CFM-ID: A web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Res.* **2014**, *42*, W94–W99. [[CrossRef](#)] [[PubMed](#)]
73. Kind, T.; Liu, K.-H.; Lee, D.Y.; DeFelice, B.; Meissen, J.K.; Fiehn, O. LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat. Methods* **2013**, *10*, 755–758. [[CrossRef](#)] [[PubMed](#)]
74. Kind, T.; Okazaki, Y.; Saito, K.; Fiehn, O. LipidBlast templates as flexible tools for creating new in-silico tandem mass spectral libraries. *Anal. Chem.* **2014**, *86*, 11024–11027. [[CrossRef](#)] [[PubMed](#)]
75. Spackman, P.R.; Bohman, B.; Karton, A.; Jayatilaka, D. Quantum chemical electron impact mass spectrum prediction for de novo structure elucidation: Assessment against experimental reference data and comparison to competitive fragmentation modeling. *Int. J. Quantum Chem.* **2017**. [[CrossRef](#)]
76. Dral, P.O.; Wu, X.; Spörkel, L.; Koslowski, A.; Thiel, W. Semiempirical quantum-chemical orthogonalization-corrected methods: Benchmarks for ground-state properties. *J. Chem. Theory Comput.* **2016**, *12*, 1097–1120. [[CrossRef](#)] [[PubMed](#)]
77. Ásgeirsson, V.; Bauer, C.A.; Grimme, S. Quantum chemical calculation of electron ionization mass spectra for general organic and inorganic molecules. *Chem. Sci.* **2017**, *8*, 4879–4895. [[CrossRef](#)] [[PubMed](#)]
78. Cautereels, J.; Claeys, M.; Geldof, D.; Blockhuys, F. Quantum chemical mass spectrometry: *ab initio* prediction of electron ionization mass spectra and identification of new fragmentation pathways. *J. Mass Spectrom.* **2016**, *51*, 602–614. [[CrossRef](#)] [[PubMed](#)]

79. Aguirre, N.F.; Díaz-Tendero, S.; Hervieux, P.-A.; Alcamí, M.; Martín, F. M3C: A Computational Approach to Describe Statistical Fragmentation of Excited Molecules and Clusters. *J. Chem. Theory Comput.* **2017**, *13*, 992–1009. [[CrossRef](#)] [[PubMed](#)]
80. Pracht, P.; Bauer, C.A.; Grimme, S. Automated and efficient quantum chemical determination and energetic ranking of molecular protonation sites. *J. Comput. Chem.* **2017**, *38*, 2618–2631. [[CrossRef](#)] [[PubMed](#)]
81. Janesko, B.G.; Li, L.; Mensing, R. Quantum Chemical Fragment Precursor Tests: Accelerating de novo annotation of tandem mass spectra. *Anal. Chim. Acta* **2017**, *995*, 52–64. [[CrossRef](#)] [[PubMed](#)]
82. Böcker, S. Searching molecular structure databases using tandem MS data: Are we there yet? *Curr. Opin. Chem. Biol.* **2017**, *36*, 1–6. [[CrossRef](#)] [[PubMed](#)]
83. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; et al. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213. [[CrossRef](#)] [[PubMed](#)]
84. Aguilar-Mogas, A.; Sales-Pardo, M.; Navarro, M.; Guimerà, R.; Yanes, O. imet: A network-based computational tool to assist in the annotation of metabolites from tandem mass spectra. *Anal. Chem.* **2017**, *89*, 3474–3482. [[CrossRef](#)] [[PubMed](#)]
85. Ridder, L.; van der Hooft, J.J.; Verhoeven, S. Automatic compound annotation from mass spectrometry data using MAGMa. *Mass Spectrom.* **2014**, *3*, S0033. [[CrossRef](#)] [[PubMed](#)]
86. Wang, Y.; Kora, G.; Bowen, B.P.; Pan, C. MIDAS: A database-searching algorithm for metabolite identification in metabolomics. *Anal. Chem.* **2014**, *86*, 9496–9503. [[CrossRef](#)] [[PubMed](#)]
87. Wang, Y.; Wang, X.; Zeng, X. MIDAS-G: A computational platform for investigating fragmentation rules of tandem mass spectrometry in metabolomics. *Metabolomics* **2017**, *13*, 116. [[CrossRef](#)]
88. Scheubert, K.; Hufsky, F.; Böcker, S. Computational mass spectrometry for small molecules. *J. Cheminform.* **2013**, *5*, 12. [[CrossRef](#)] [[PubMed](#)]
89. Hufsky, F.; Scheubert, K.; Böcker, S. Computational mass spectrometry for small-molecule fragmentation. *TrAC Trends Anal. Chem.* **2014**, *53*, 41–48. [[CrossRef](#)]
90. Hufsky, F.; Scheubert, K.; Böcker, S. New kids on the block: Novel informatics methods for natural product discovery. *Nat. Prod. Rep.* **2014**, *31*, 807–817. [[CrossRef](#)] [[PubMed](#)]
91. Hufsky, F.; Böcker, S. Mining molecular structure databases: Identification of small molecules based on fragmentation mass spectrometry data. *Mass Spectrom. Rev.* **2017**, *36*, 624–633. [[CrossRef](#)] [[PubMed](#)]
92. Ruttkies, C.; Schymanski, E.L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: Incorporating strategies beyond in silico fragmentation. *J. Cheminform.* **2016**, *8*, 3. [[CrossRef](#)] [[PubMed](#)]
93. Gerlich, M.; Neumann, S. MetFusion: Integration of compound identification strategies. *J. Mass Spectrom.* **2013**, *48*, 291–298. [[CrossRef](#)] [[PubMed](#)]
94. Witting, M.; Ruttkies, C.; Neumann, S.; Schmitt-Kopplin, P. LipidFrag: Improving reliability of in silico fragmentation of lipids and application to the *Caenorhabditis elegans* lipidome. *PLoS ONE* **2017**, *12*, e0172311. [[CrossRef](#)] [[PubMed](#)]
95. Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinform.* **2007**, *8*, 105. [[CrossRef](#)] [[PubMed](#)]
96. Vaniya, A.; Samra, S.N.; Palazoglu, M.; Tsugawa, H.; Fiehn, O. Using MS-FINDER for identifying 19 natural products in the CASMI 2016 contest. *Phytochem. Lett.* **2017**, *21*, 306–312. [[CrossRef](#)]
97. Tsugawa, H.; Kind, T.; Nakabayashi, R.; Yukihiro, D.; Tanaka, W.; Cajka, T.; Saito, K.; Fiehn, O.; Arita, M. Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. *Anal. Chem.* **2016**, *88*, 7946–7958. [[CrossRef](#)] [[PubMed](#)]
98. Duhrkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Bocker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12580–12585. [[CrossRef](#)] [[PubMed](#)]
99. Heinonen, M.; Shen, H.; Zamboni, N.; Rousu, J. Metabolite identification and molecular fingerprint prediction through machine learning. *Bioinformatics* **2012**, *28*, 2333–2341. [[CrossRef](#)] [[PubMed](#)]
100. Brouard, C.; Shen, H.; Duhrkop, K.; d’Alché-Buc, F.; Böcker, S.; Rousu, J. Fast metabolite identification with input output kernel regression. *Bioinformatics* **2016**, *32*, i28–i36. [[CrossRef](#)] [[PubMed](#)]
101. Bocker, S.; Letzel, M.C.; Liptak, Z.; Pervukhin, A. SIRIUS: Decomposing isotope patterns for metabolite identification. *Bioinformatics* **2009**, *25*, 218–224. [[CrossRef](#)] [[PubMed](#)]

102. Rasche, F.; Svatoš, A.; Maddula, R.K.; Böttcher, C.; Böcker, S. Computing fragmentation trees from tandem mass spectrometry data. *Anal. Chem.* **2010**, *83*, 1243–1251. [[CrossRef](#)] [[PubMed](#)]
103. Laponogov, I.; Sadawi, N.; Galea, D.; Mirnezami, R.; Veselkov, K.A.; Wren, J. ChemDistiller: An engine for metabolite annotation in mass spectrometry. *Bioinformatics* **2018**, *1*, 7. [[CrossRef](#)] [[PubMed](#)]
104. Kaufmann, A.; Butcher, P.; Maden, K.; Walker, S.; Widmer, M. Using In Silico Fragmentation to Improve Routine Residue Screening in Complex Matrices. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 2705–2715. [[CrossRef](#)] [[PubMed](#)]
105. Blazenovic, I.; Kind, T.; Torbasinovic, H.; Obrenovic, S.; Mehta, S.S.; Tsugawa, H.; Wermuth, T.; Schauer, N.; Jahn, M.; Biedendieck, R.; et al. Comprehensive comparison of in silico MS/MS fragmentation tools of the CASMI contest: Database boosting is needed to achieve 93% accuracy. *J. Cheminform.* **2017**, *9*, 32. [[CrossRef](#)] [[PubMed](#)]
106. Schymanski, E.L.; Ruttkies, C.; Krauss, M.; Brouard, C.; Kind, T.; Dührkop, K.; Allen, F.; Vaniya, A.; Verdegem, D.; Böcker, S.; et al. Critical Assessment of Small Molecule Identification 2016: Automated methods. *J. Cheminform.* **2017**, *9*, 22. [[CrossRef](#)] [[PubMed](#)]
107. Kaliszan, R. QSRR: Quantitative structure-(chromatographic) retention relationships. *Chem. Rev.* **2007**, *107*, 3212–3246. [[CrossRef](#)] [[PubMed](#)]
108. Stein, S.E.; Babushok, V.I.; Brown, R.L.; Linstrom, P.J. Estimation of Kovats retention indices using group contributions. *J. Chem. Inf. Model.* **2007**, *47*, 975–980. [[CrossRef](#)] [[PubMed](#)]
109. Navarro-Reig, M.; Ortiz-Villanueva, E.; Tauler, R.; Jaumot, J. Modelling of Hydrophilic Interaction Liquid Chromatography Stationary Phases Using Chemometric Approaches. *Metabolites* **2017**, *7*, 54. [[CrossRef](#)] [[PubMed](#)]
110. Sahigara, F.; Mansouri, K.; Ballabio, D.; Mauri, A.; Consonni, V.; Todeschini, R. Comparison of different approaches to define the applicability domain of QSAR models. *Molecules* **2012**, *17*, 4791–4810. [[CrossRef](#)] [[PubMed](#)]
111. Hall, L.M.; Hill, D.W.; Bugden, K.; Cawley, S.; Hall, L.H.; Chen, M.-H.; Grant, D.F. Development of a Reverse Phase HPLC Retention Index Model for Nontargeted Metabolomics Using Synthetic Compounds. *J. Chem. Inf. Model.* **2018**, *58*, 591–604. [[CrossRef](#)] [[PubMed](#)]
112. Barnes, B.B.; Wilson, M.B.; Carr, P.W.; Vitha, M.F.; Broeckling, C.D.; Heuberger, A.L.; Prenni, J.; Janis, G.C.; Corcoran, H.; Snow, N.H. “Retention projection” enables reliable use of shared gas chromatographic retention data across laboratories, instruments, and methods. *Anal. Chem.* **2013**, *85*, 11650–11657. [[CrossRef](#)] [[PubMed](#)]
113. Stanstrup, J.; Neumann, S.; Vrhovsek, U. PredRet: Prediction of retention time by direct mapping between multiple chromatographic systems. *Anal. Chem.* **2015**, *87*, 9421–9428. [[CrossRef](#)] [[PubMed](#)]
114. Boswell, P.G.; Abate-Pella, D.; Hewitt, J.T. Calculation of retention time tolerance windows with absolute confidence from shared liquid chromatographic retention data. *J. Chromatogr. A* **2015**, *1412*, 52–58. [[CrossRef](#)] [[PubMed](#)]
115. Mauri, A.; Consonni, V.; Pavan, M.; Todeschini, R. Dragon software: An easy approach to molecular descriptor calculations. *Match* **2006**, *56*, 237–248.
116. Hong, H.; Xie, Q.; Ge, W.; Qian, F.; Fang, H.; Shi, L.; Su, Z.; Perkins, R.; Tong, W. Mold2, molecular descriptors from 2D structures for chemoinformatics and toxicoinformatics. *J. Chem. Inf. Model.* **2008**, *48*, 1337–1344. [[CrossRef](#)] [[PubMed](#)]
117. Yap, C.W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466–1474. [[CrossRef](#)] [[PubMed](#)]
118. Friedrich, N.-O.; de Bruyn Kops, C.; Flachsenberg, F.; Sommer, K.; Rarey, M.; Kirchmair, J. Benchmarking Commercial Conformer Ensemble Generators. *J. Chem. Inf. Model.* **2017**, *57*, 2719–2728. [[CrossRef](#)] [[PubMed](#)]
119. Kanal, I.Y.; Keith, J.A.; Hutchison, G.R. A sobering assessment of small-molecule force field methods for low energy conformer predictions. *Int. J. Quantum Chem.* **2018**, *118*, e25512. [[CrossRef](#)]
120. Ma, J.; Sheridan, R.P.; Liaw, A.; Dahl, G.E.; Svetnik, V. Deep neural nets as a method for quantitative structure–activity relationships. *J. Chem. Inf. Model.* **2015**, *55*, 263–274. [[CrossRef](#)] [[PubMed](#)]
121. Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we need hundreds of classifiers to solve real world classification problems. *J. Mach. Learn. Res.* **2014**, *15*, 3133–3181.
122. He, H.; Zhang, W.; Zhang, S. A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Syst. Appl.* **2018**, *98*, 105–117. [[CrossRef](#)]

123. Falchi, F.; Bertozzi, S.M.; Ottonello, G.; Ruda, G.F.; Colombano, G.; Fiorelli, C.; Martucci, C.; Bertorelli, R.; Scarpelli, R.; Cavalli, A. Kernel-based, partial least squares quantitative structure-retention relationship model for UPLC retention time prediction: A useful tool for metabolite identification. *Anal. Chem.* **2016**, *88*, 9510–9517. [[CrossRef](#)] [[PubMed](#)]
124. Taraji, M.; Haddad, P.R.; Amos, R.I.; Talebi, M.; Szucs, R.; Dolan, J.W.; Pohl, C.A. Use of dual-filtering to create training sets leading to improved accuracy in quantitative structure-retention relationships modelling for hydrophilic interaction liquid chromatographic systems. *J. Chromatogr. A* **2017**, *1507*, 53–62. [[CrossRef](#)] [[PubMed](#)]
125. Taraji, M.; Haddad, P.R.; Amos, R.I.; Talebi, M.; Szucs, R.; Dolan, J.W.; Pohl, C.A. Prediction of retention in hydrophilic interaction liquid chromatography using solute molecular descriptors based on chemical structures. *J. Chromatogr. A* **2017**, *1486*, 59–67. [[CrossRef](#)] [[PubMed](#)]
126. Bruderer, T.; Varesio, E.; Hopfgartner, G. The use of LC predicted retention times to extend metabolites identification with SWATH data acquisition. *J. Chromatogr. B* **2017**, *1071*, 3–10. [[CrossRef](#)] [[PubMed](#)]
127. Creek, D.J.; Jankevics, A.; Breitling, R.; Watson, D.G.; Barrett, M.P.; Burgess, K.E. Toward global metabolomics analysis with hydrophilic interaction liquid chromatography–mass spectrometry: Improved metabolite identification by retention time prediction. *Anal. Chem.* **2011**, *83*, 8703–8710. [[CrossRef](#)] [[PubMed](#)]
128. Aalizadeh, R.; Thomaidis, N.S.; Bletsou, A.A.; Gago-Ferrero, P. Quantitative Structure–Retention Relationship Models to Support Nontarget High-Resolution Mass Spectrometric Screening of Emerging Contaminants in Environmental Samples. *J. Chem. Inf. Model.* **2016**, *56*, 1384–1398. [[CrossRef](#)] [[PubMed](#)]
129. Aicheler, F.; Li, J.; Hoene, M.; Lehmann, R.; Xu, G.; Kohlbacher, O. Retention time prediction improves identification in nontargeted lipidomics approaches. *Anal. Chem.* **2015**, *87*, 7698–7704. [[CrossRef](#)] [[PubMed](#)]
130. Wolfer, A.M.; Lozano, S.; Umbdenstock, T.; Croixmarie, V.; Arrault, A.; Vayer, P. UPLC–MS retention time prediction: A machine learning approach to metabolite identification in untargeted profiling. *Metabolomics* **2016**, *12*, 8. [[CrossRef](#)]
131. Cao, M.; Fraser, K.; Huege, J.; Featonby, T.; Rasmussen, S.; Jones, C. Predicting retention time in hydrophilic interaction liquid chromatography mass spectrometry and its use for peak annotation in metabolomics. *Metabolomics* **2015**, *11*, 696–706. [[CrossRef](#)] [[PubMed](#)]
132. Møllerup, C.B.; Mardal, M.; Dalsgaard, P.W.; Linnet, K.; Barron, L.P. Prediction of collision cross section and retention time for broad scope screening in gradient reversed-phase liquid chromatography-ion mobility-high resolution accurate mass spectrometry. *J. Chromatogr. A* **2018**, *1542*, 82–88. [[CrossRef](#)] [[PubMed](#)]
133. Hall, L.M.; Hall, L.H.; Kertesz, T.M.; Hill, D.W.; Sharp, T.R.; Oblak, E.Z.; Dong, Y.W.; Wishart, D.S.; Chen, M.H.; Grant, D.F. Development of Ecom₅₀ and retention index models for nontargeted metabolomics: Identification of 1,3-dicyclohexylurea in human serum by HPLC/mass spectrometry. *J. Chem. Inf. Model.* **2012**, *52*, 1222–1237. [[CrossRef](#)] [[PubMed](#)]
134. Eugster, P.J.; Boccard, J.; Debrus, B.; Bréant, L.; Wolfender, J.-L.; Martel, S.; Carrupt, P.-A. Retention time prediction for dereplication of natural products (C_xH_yO_z) in LC–MS metabolite profiling. *Phytochemistry* **2014**, *108*, 196–207. [[CrossRef](#)] [[PubMed](#)]
135. Chouinard, C.D.; Cruzeiro, V.W.D.; Beekman, C.R.; Roitberg, A.E.; Yost, R.A. Investigating Differences in Gas-Phase Conformations of 25-Hydroxyvitamin D3 Sodiater Epimers using Ion Mobility-Mass Spectrometry and Theoretical Modeling. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1497–1505. [[CrossRef](#)] [[PubMed](#)]
136. May, J.C.; Goodwin, C.R.; Lareau, N.M.; Leaptrot, K.L.; Morris, C.B.; Kurulugama, R.T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E. Conformational ordering of biomolecules in the gas phase: Nitrogen collision cross sections measured on a prototype high resolution drift tube ion mobility-mass spectrometer. *Anal. Chem.* **2014**, *86*, 2107–2116. [[CrossRef](#)] [[PubMed](#)]
137. May, J.C.; McLean, J.A. Ion mobility-mass spectrometry: Time-dispersive instrumentation. *Anal. Chem.* **2015**, *87*, 1422–1436. [[CrossRef](#)] [[PubMed](#)]
138. D’Atri, V.; Causon, T.; Hernandez-Alba, O.; Mutabazi, A.; Veuthey, J.L.; Cianferani, S.; Guilleme, D. Adding a new separation dimension to MS and LC–MS: What is the utility of ion mobility spectrometry? *J. Sep. Sci.* **2018**, *41*, 20–67. [[CrossRef](#)] [[PubMed](#)]

139. Zheng, X.; Aly, N.A.; Zhou, Y.; Dupuis, K.T.; Bilbao, A.; Paurus, V.L.; Orton, D.J.; Wilson, R.; Payne, S.H.; Smith, R.D.; et al. A structural examination and collision cross section database for over 500 metabolites and xenobiotics using drift tube ion mobility spectrometry. *Chem. Sci.* **2017**, *8*, 7724–7736. [[CrossRef](#)] [[PubMed](#)]
140. Ma, J.; Casey, C.P.; Zheng, X.; Ibrahim, Y.M.; Wilkins, C.S.; Renslow, R.S.; Thomas, D.G.; Payne, S.H.; Monroe, M.E.; Smith, R.D.; et al. PIXiE: An algorithm for automated ion mobility arrival time extraction and collision cross section calculation using global data association. *Bioinformatics* **2017**, *33*, 2715–2722. [[CrossRef](#)] [[PubMed](#)]
141. Ewing, M.A.; Glover, M.S.; Clemmer, D.E. Hybrid ion mobility and mass spectrometry as a separation tool. *J. Chromatogr. A* **2016**, *1439*, 3–25. [[CrossRef](#)] [[PubMed](#)]
142. May, J.C.; McLean, J.A. Advanced multidimensional separations in mass spectrometry: Navigating the big data deluge. *Annu. Rev. Anal. Chem.* **2016**, *9*, 387–409. [[CrossRef](#)] [[PubMed](#)]
143. Lapthorn, C.; Pullen, F.; Chowdhry, B.Z. Ion mobility spectrometry-mass spectrometry (IMS-MS) of small molecules: Separating and assigning structures to ions. *Mass Spectrom. Rev.* **2013**, *32*, 43–71. [[CrossRef](#)] [[PubMed](#)]
144. Canterbury, J.D.; Yi, X.; Hoopmann, M.R.; MacCoss, M.J. Assessing the dynamic range and peak capacity of nanoflow LC–FAIMS–MS on an ion trap mass spectrometer for proteomics. *Anal. Chem.* **2008**, *80*, 6888–6897. [[CrossRef](#)] [[PubMed](#)]
145. Hines, K.M.; Ross, D.H.; Davidson, K.L.; Bush, M.F.; Xu, L. Large-scale structural characterization of drug and drug-like compounds by high-throughput ion mobility-mass spectrometry. *Anal. Chem.* **2017**, *89*, 9023–9030. [[CrossRef](#)] [[PubMed](#)]
146. Nichols, C.M.; May, J.C.; Sherrod, S.D.; McLean, J.A. Automated flow injection method for the high precision determination of drift tube ion mobility collision cross sections. *Analyst* **2018**, *143*, 1556–1559. [[CrossRef](#)] [[PubMed](#)]
147. Mairinger, T.; Causon, T.J.; Hann, S. The potential of ion mobility–mass spectrometry for non-targeted metabolomics. *Curr. Opin. Chem. Biol.* **2018**, *42*, 9–15. [[CrossRef](#)] [[PubMed](#)]
148. Metz, T.O.; Baker, E.S.; Schymanski, E.L.; Renslow, R.S.; Thomas, D.G.; Causon, T.J.; Webb, I.K.; Hann, S.; Smith, R.D.; Teeguarden, J.G. Integrating ion mobility spectrometry into mass spectrometry-based exposome measurements: What can it add and how far can it go? *Bioanalysis* **2017**, *9*, 81–98. [[CrossRef](#)] [[PubMed](#)]
149. Lapthorn, C.; Pullen, F.S.; Chowdhry, B.Z.; Wright, P.; Perkins, G.L.; Heredia, Y. How useful is molecular modelling in combination with ion mobility mass spectrometry for ‘small molecule’ ion mobility collision cross-sections? *Analyst* **2015**, *140*, 6814–6823. [[CrossRef](#)] [[PubMed](#)]
150. Zhou, Z.; Tu, J.; Zhu, Z.-J. Advancing the large-scale CCS database for metabolomics and lipidomics at the machine-learning era. *Curr. Opin. Chem. Biol.* **2018**, *42*, 34–41. [[CrossRef](#)] [[PubMed](#)]
151. Bijlsma, L.; Bade, R.; Celma, A.; Mullin, L.; Cleland, G.; Stead, S.; Hernandez, F.; Sancho, J.V. Prediction of collision cross-section values for small molecules: Application to pesticide residue analysis. *Anal. Chem.* **2017**, *89*, 6583–6589. [[CrossRef](#)] [[PubMed](#)]
152. Zhou, Z.; Shen, X.; Tu, J.; Zhu, Z.-J. Large-scale prediction of collision cross-section values for metabolites in ion mobility-mass spectrometry. *Anal. Chem.* **2016**, *88*, 11084–11091. [[CrossRef](#)] [[PubMed](#)]
153. Kyle, J.E.; Zhang, X.; Weitz, K.K.; Monroe, M.E.; Ibrahim, Y.M.; Moore, R.J.; Cha, J.; Sun, X.; Lovelace, E.S.; Wagoner, J. Uncovering biologically significant lipid isomers with liquid chromatography, ion mobility spectrometry and mass spectrometry. *Analyst* **2016**, *141*, 1649–1659. [[CrossRef](#)] [[PubMed](#)]
154. Zhou, Z.; Tu, J.; Xiong, X.; Shen, X.; Zhu, Z.-J. LipidCCS: Prediction of Collision Cross-Section Values for Lipids with High Precision to Support Ion Mobility–Mass Spectrometry-Based Lipidomics. *Anal. Chem.* **2017**, *89*, 9559–9566. [[CrossRef](#)] [[PubMed](#)]
155. Paglia, G.; Williams, J.P.; Menikarachchi, L.; Thompson, J.W.; Tyldesley-Worster, R.; Halldórsson, S.; Rolfsson, O.; Moseley, A.; Grant, D.; Langridge, J.; et al. Ion mobility derived collision cross sections to support metabolomics applications. *Anal. Chem.* **2014**, *86*, 3985–3993. [[CrossRef](#)] [[PubMed](#)]
156. Hernández-Mesa, M.; Le Bizec, B.; Monteau, F.; García-Campaña, A.M.; Dervilly-Pinel, G. Collision Cross Section (CCS) database: An additional measure to characterize steroids. *Anal. Chem.* **2018**, *90*, 4616–4625. [[CrossRef](#)] [[PubMed](#)]

157. Zheng, X.; Dupuis, K.T.; Aly, N.A.; Zhou, Y.; Smith, F.B.; Tang, K.; Smith, R.D.; Baker, E.S. Utilizing ion mobility spectrometry and mass spectrometry for the analysis of polycyclic aromatic hydrocarbons, polychlorinated biphenyls, polybrominated diphenyl ethers and their metabolites. *Anal. Chim. Acta* **2018**. [[CrossRef](#)]
158. Gabelica, V.; Marklund, E. Fundamentals of ion mobility spectrometry. *Curr. Opin. Chem. Biol.* **2018**, *42*, 51–59. [[CrossRef](#)] [[PubMed](#)]
159. Wyttenbach, T.; Pierson, N.A.; Clemmer, D.E.; Bowers, M.T. Ion mobility analysis of molecular dynamics. *Annu. Rev. Phys. Chem.* **2014**, *65*, 175–196. [[CrossRef](#)] [[PubMed](#)]
160. Warnke, S.; Seo, J.; Boschmans, J.; Sobott, F.; Scrivens, J.H.; Bleiholder, C.; Bowers, M.T.; Gewinner, S.; Schöllkopf, W.; Pagel, K. Protomers of benzocaine: Solvent and permittivity dependence. *J. Am. Chem. Soc.* **2015**, *137*, 4236–4242. [[CrossRef](#)] [[PubMed](#)]
161. Laphorn, C.; Dines, T.J.; Chowdhry, B.Z.; Perkins, G.L.; Pullen, F.S. Can ion mobility mass spectrometry and density functional theory help elucidate protonation sites in 'small' molecules? *Rapid Commun. Mass Spectrom.* **2013**, *27*, 2399–2410. [[CrossRef](#)] [[PubMed](#)]
162. Boschmans, J.; Jacobs, S.; Williams, J.P.; Palmer, M.; Richardson, K.; Giles, K.; Laphorn, C.; Herrebout, W.A.; Lemi re, F.; Sobott, F. Combining density functional theory (DFT) and collision cross-section (CCS) calculations to analyze the gas-phase behaviour of small molecules and their protonation site isomers. *Analyst* **2016**, *141*, 4044–4054. [[CrossRef](#)] [[PubMed](#)]
163. Stow, S.M.; Causon, T.J.; Zheng, X.; Kurulugama, R.T.; Mairinger, T.; May, J.C.; Rennie, E.E.; Baker, E.S.; Smith, R.D.; McLean, J.A. An interlaboratory evaluation of drift tube ion mobility–mass spectrometry collision cross section measurements. *Anal. Chem.* **2017**, *89*, 9048–9055. [[CrossRef](#)] [[PubMed](#)]
164. Zhou, Z.; Xiong, X.; Zhu, Z.-J. MetCCS predictor: A web server for predicting collision cross-section values of metabolites in ion mobility-mass spectrometry based metabolomics. *Bioinformatics* **2017**, *33*, 2235–2237. [[CrossRef](#)] [[PubMed](#)]
165. Ibrahim, Y.M.; Hamid, A.M.; Cox, J.T.; Garimella, S.V.; Smith, R.D. Ion Elevators and Escalators in Multilevel Structures for Lossless Ion Manipulations. *Anal. Chem.* **2017**, *89*, 1972–1977. [[CrossRef](#)] [[PubMed](#)]
166. Ibrahim, Y.M.; Hamid, A.M.; Deng, L.; Garimella, S.V.; Webb, I.K.; Baker, E.S.; Smith, R.D. New frontiers for mass spectrometry based upon structures for lossless ion manipulations. *Analyst* **2017**, *142*, 1010–1021. [[CrossRef](#)] [[PubMed](#)]
167. Weber, R.J.M.; Lawson, T.N.; Salek, R.M.; Ebbels, T.M.D.; Glen, R.C.; Goodacre, R.; Griffin, J.L.; Haug, K.; Koulman, A.; Moreno, P.; et al. Computational tools and workflows in metabolomics: An international survey highlights the opportunity for harmonisation through Galaxy. *Metabolomics* **2017**, *13*, 12. [[CrossRef](#)] [[PubMed](#)]
168. Allard, P.-M.; P resse, T.; Bisson, J.; Gindro, K.; Marcourt, L.; Pham, V.C.; Roussi, F.; Litaudon, M.; Wolfender, J.-L. Integration of molecular networking and in-silico MS/MS fragmentation for natural products dereplication. *Anal. Chem.* **2016**, *88*, 3317–3323. [[CrossRef](#)] [[PubMed](#)]
169. Allard, P.-M.; Genta-Jouve, G.; Wolfender, J.-L. Deep metabolome annotation in natural products research: Towards a virtuous cycle in metabolite identification. *Curr. Opin. Chem. Biol.* **2017**, *36*, 40–49. [[CrossRef](#)] [[PubMed](#)]
170. De la Fuente, A.G.; Godzien, J.; L pez, M.F.; Rup rez, F.J.; Barbas, C.; Otero, A. Knowledge-based metabolite annotation tool: CEU Mass Mediator. *J. Pharm. Biomed. Anal.* **2018**, *154*, 138–149. [[CrossRef](#)] [[PubMed](#)]
171. Broeckling, C.D.; Ganna, A.; Layer, M.; Brown, K.; Sutton, B.; Ingelsson, E.; Peers, G.; Prenni, J.E. Enabling Efficient and Confident Annotation of LC–MS Metabolomics Data through MS1 Spectrum and Time Prediction. *Anal. Chem.* **2016**, *88*, 9226–9234. [[CrossRef](#)] [[PubMed](#)]
172. Uppal, K.; Walker, D.I.; Jones, D.P. xMSannotator: An R package for network-based annotation of high-resolution metabolomics data. *Anal. Chem.* **2017**, *89*, 1063–1067. [[CrossRef](#)] [[PubMed](#)]
173. Van Der Hooft, J.J.J.; Wandy, J.; Barrett, M.P.; Burgess, K.E.; Rogers, S. Topic modeling for untargeted substructure exploration in metabolomics. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 13738–13743. [[CrossRef](#)] [[PubMed](#)]
174. Naz, S.; Gallart-Ayala, H.; Reinke, S.N.; Mathon, C.; Blankley, R.; Chaleckis, R.; Wheelock, C.E. Development of a Liquid Chromatography–High Resolution Mass Spectrometry Metabolomics Method with High Specificity for Metabolite Identification Using All Ion Fragmentation Acquisition. *Anal. Chem.* **2017**, *89*, 7933–7942. [[CrossRef](#)] [[PubMed](#)]

175. Hu, M.; Müller, E.; Schymanski, E.L.; Ruttkies, C.; Schulze, T.; Brack, W.; Krauss, M. Performance of combined fragmentation and retention prediction for the identification of organic micropollutants by LC-HRMS. *Anal. Bioanal. Chem.* **2018**, *410*, 1931–1941. [[CrossRef](#)] [[PubMed](#)]
176. Jaeger, C.; Méret, M.; Schmitt, C.A.; Lisec, J. Compound annotation in liquid chromatography/high-resolution mass spectrometry based metabolomics: Robust adduct ion determination as a prerequisite to structure prediction in electrospray ionization mass spectra. *Rapid Commun. Mass Spectrom.* **2017**, *31*, 1261–1266. [[CrossRef](#)] [[PubMed](#)]
177. Zani, C.L.; Carroll, A.R. Database for rapid dereplication of known natural products using data from MS and fast NMR experiments. *J. Nat. Prod.* **2017**, *80*, 1758–1766. [[CrossRef](#)] [[PubMed](#)]
178. De Vijlder, T.; Valkenburg, D.; Lemièr, F.; Romijn, E.P.; Laukens, K.; Cuyckens, F. A tutorial in small molecule identification via electrospray ionization-mass spectrometry: The practical art of structural elucidation. *Mass Spectrom. Rev.* **2017**. [[CrossRef](#)] [[PubMed](#)]
179. Werner, E.; Heilier, J.-F.; Ducruix, C.; Ezan, E.; Junot, C.; Tabet, J.-C. Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends. *J. Chromatogr. B* **2008**, *871*, 143–163. [[CrossRef](#)] [[PubMed](#)]
180. Fitzgerald, B.L.; Mahapatra, S.; Farmer, D.K.; McNeil, M.R.; Casero, R.A., Jr.; Belisle, J.T. Elucidating the Structure of N^1 -Acetylisoputrescine: A Novel Polyamine Catabolite in Human Urine. *ACS Omega* **2017**, *2*, 3921–3930. [[CrossRef](#)] [[PubMed](#)]
181. Cajka, T.; Garay, L.A.; Sitepu, I.R.; Boundy-Mills, K.L.; Fiehn, O. Multiplatform mass spectrometry-based approach identifies extracellular glycolipids of the yeast *Rhodotorula babjevae* UCDFST 04-877. *J. Nat. Prod.* **2016**, *79*, 2580–2589. [[CrossRef](#)] [[PubMed](#)]
182. Nikolić, D. CASMI 2016: A manual approach for dereplication of natural products using tandem mass spectrometry. *Phytochem. Lett.* **2017**, *21*, 292–296. [[CrossRef](#)] [[PubMed](#)]
183. Ruttkies, C.; Gerlich, M.; Neumann, S. Tackling CASMI 2012: Solutions from MetFrag and MetFusion. *Metabolites* **2013**, *3*, 623–636. [[CrossRef](#)] [[PubMed](#)]
184. Nishioka, T.; Kasama, T.; Kinumi, T.; Makabe, H.; Matsuda, F.; Miura, D.; Miyashita, M.; Nakamura, T.; Tanaka, K.; Yamamoto, A. Winners of CASMI2013: Automated tools and challenge data. *Mass Spectrom.* **2014**, *3*, S0039. [[CrossRef](#)] [[PubMed](#)]
185. Newsome, A.G.; Nikolic, D. CASMI 2013: Identification of small molecules by tandem mass spectrometry combined with database and literature mining. *Mass Spectrom.* **2014**, *3*, S0034. [[CrossRef](#)] [[PubMed](#)]
186. Dührkop, K.; Scheubert, K.; Böcker, S. Molecular formula identification with SIRIUS. *Metabolites* **2013**, *3*, 506–516. [[CrossRef](#)] [[PubMed](#)]
187. Shen, H.; Zamboni, N.; Heinonen, M.; Rousu, J. Metabolite identification through machine learning—Tackling casmi challenge using FingerID. *Metabolites* **2013**, *3*, 484–505. [[CrossRef](#)] [[PubMed](#)]
188. Gewin, V. Data sharing: An open mind on open data. *Nature* **2016**, *529*, 117–119. [[CrossRef](#)] [[PubMed](#)]
189. Spicer, R.A.; Steinbeck, C. A lost opportunity for science: Journals promote data sharing in metabolomics but do not enforce it. *Metabolomics* **2018**, *14*, 16. [[CrossRef](#)] [[PubMed](#)]
190. Rübel, O.; Greiner, A.; Cholia, S.; Louie, K.; Bethel, E.W.; Northen, T.R.; Bowen, B.P. OpenMSI: A high-performance web-based platform for mass spectrometry imaging. *Anal. Chem.* **2013**, *85*, 10354–10361. [[CrossRef](#)] [[PubMed](#)]
191. Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; et al. ClassyFire: Automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminform.* **2016**, *8*, 61. [[CrossRef](#)] [[PubMed](#)]

